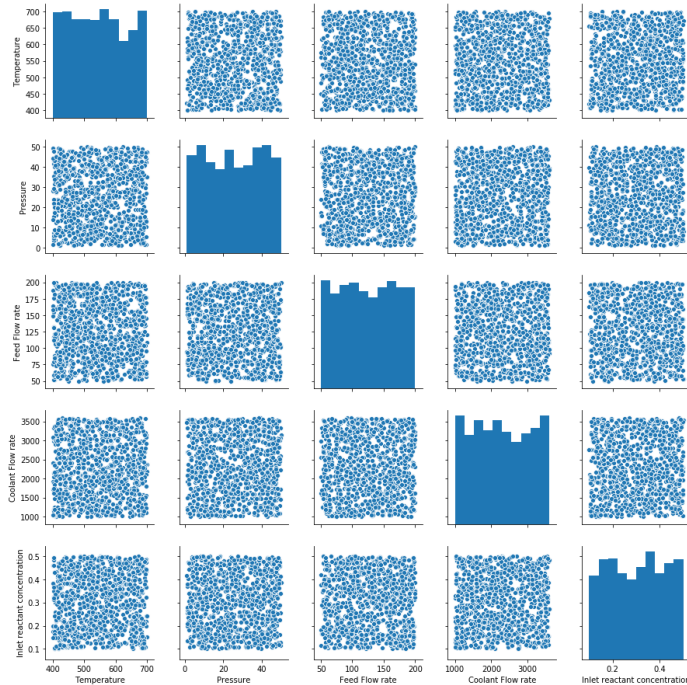


## Question 2:-

**Python 3** was used to analyse the data as well as for creating the Logistic Regression model.

Given data => 1000 samples ( 5 attributes, 1 label)

Scatter plot between various attributes confirms the absence of linear dependency among the attributes.



The labels were encoded as:

1 → Pass

0 → Fail

Now, the 5 attributes were scaled to **mean 0** and **standard deviation 1** using the formula:-

$$z = \frac{x - \mu}{s}$$

The scaled data was fitted to the **Logit** function (labels vs attributes).

	Coef.	Std.Err.	z	P> z	[0.025	0.975]
Temperature	-0.0013	0.0012	-1.0945	0.2737	-0.0036	0.0010
Pressure	-0.0274	0.0074	-3.6884	0.0002	-0.0419	-0.0128
Feed Flow rate	-0.0116	0.0024	-4.7534	0.0000	-0.0164	-0.0068
Coolant Flow rate	0.0039	0.0002	16.5989	0.0000	0.0034	0.0043
Inlet reactant concentration	-0.6702	0.8867	-0.7558	0.4498	-2.4082	1.0678

Since, the P-values of **Temperature** and **Inlet Reactant Concentration** are quite high these features were treated as **Insignificant** i.e they do not contribute to passing or failing of the reactor .

Now, a Logistic Regression model for Binary classification was built which will take Pressure, Feed Flow rate and Coolant Flow rate as inputs and give 1/0 as outputs. (1 → Pass, 0 → Fail).

70% of the data was used for training the model and 30% for testing.

After training the model, its performance was evaluated on the test dataset.

### Results:-

The equation of the decision boundary (Hyperplane) is:-

$$h = 0.8393 - 0.0354 * (\text{Pressure}) - 0.015 * (\text{Feed Flow rate}) + 0.004 * (\text{Coolant Flow rate}) = 0$$

Where Pressure, Feed Flow Rate and Coolant Flow rate are the scaled values.

When a new datapoint needs to be classified,  $\text{sigmoid}(h)$  will be calculated.

$$\text{sigmoid}(h) = \frac{1}{1 + e^{-h}}$$

If  $\text{sigmoid}(h) > 0.5 \rightarrow 1$  (Pass)

If  $\text{sigmoid}(h) < 0.5 \rightarrow 0$  (Fail)

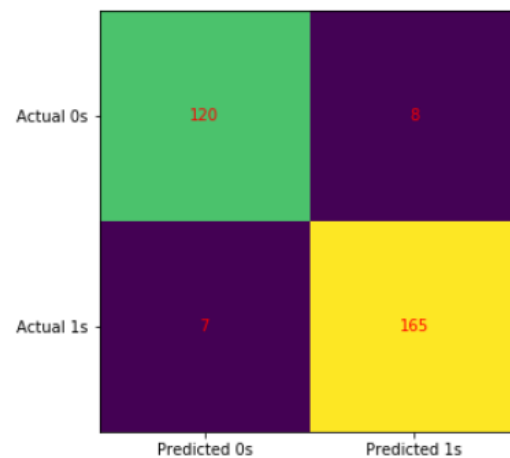


Figure:- Confusion Matrix

	precision	recall	f1-score	support
0	0.94	0.94	0.94	128
1	0.95	0.96	0.96	172
accuracy			0.95	300
macro avg	0.95	0.95	0.95	300
weighted avg	0.95	0.95	0.95	300

Figure:- Classification Report

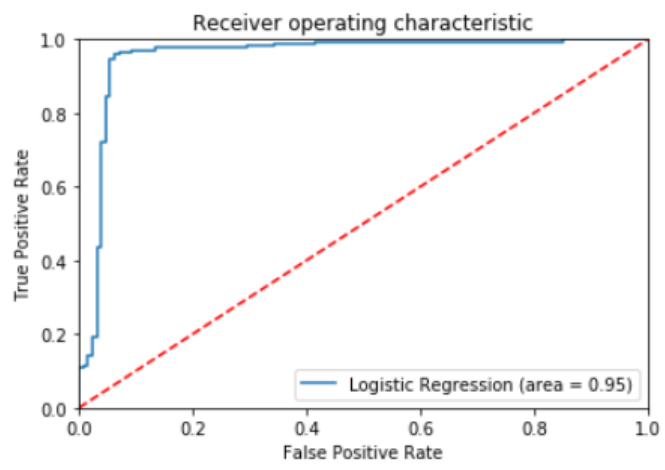


Figure:- Receiver Operating Characteristics